**FINAL VERSION (12/16/2002)**

**Climate Change Research Initiative Workshop – December 3-5, 2002**
**Breakout Group 2 – December 4, 2002; 0830-1030**
**Session 10: Climate Quality Data Management, Rapporteur's Report**
**Harding Room – Marriott Wardman Park Hotel**
**Washington, DC**

Moderator: Ms. Martha Maiden – National Aeronautics and Space Administration
Presenter: Dr. Margarita Conkright – National Oceanic and Atmospheric Administration

Panel:
Dr. Roberta Balstad Miller – Columbia University, CIESIN
Dr. Tom Boden – Oak Ridge National Laboratory
Dr. H. Lee Dantzler – National Oceanic and Atmospheric Administration
Dr. Richard Rosen – Atmospheric and Environmental Research (AER), Inc.

Rapportuers:
Mr. Howard Diamond – National Oceanic and Atmospheric Administration
Dr. Wanda Ferrell - U.S. Department of Energy

The session was called to order at 0830 hours and Ms. Martha Maiden provided a general introduction and overview to the session. Participants were reminded that any comments on the strategic plan needed to be submitted in writing to the http://climatescience.gov web site; she thanked the authors of the chapters of the strategic plan as well as the Climate Change Science Program Office for putting the workshop together, and then introduced the participants and turned over the floor to Dr. Margarita Conkright for an overview of the Climate Change Research Initiative's (CCRI) goals as they applied to the data and information across disciplines as they were documented in chapters 3 and 12 of the CCRI Strategic Plan. Dr. Conkright was one of the lead authors on chapter 3 and her intent was to provide a framework of what has taken place to help facilitate discussion of the subject.

Dr. Conkright started by speaking to the availability of the climate record as a key issue. Some areas of data management and information cover regional issues, field campaigns, and analysis of data and model outputs that require the transfer of data and information. A key challenge for data management is to provide seamless access to data and information that is required by researchers, data managers and decision makers. She then explained the data management elements outlined the CCRI's vision for Data and Information Management, which involves the following precepts:

- Collect and manage data in multiple locations, including Federal, state, and local agencies, academic institutions, non-governmental organizations, and private companies

- Enable users to find and access these data via the Internet, utilizing sophisticated systems for data search and retrieval; and

- Facilitate integrated applications related to environmental and resource management, disaster mitigation and emergency response, and other data-dependent activities.

Some of the challenges are the management of large and diverse data sets, particularly with the advent of satellite data; and the issues of storage, access, quality control, data documentation, and associated costs are major concerns

Another challenge is the preservation of the historical data and the problems associated with losing them due to such things as problems with media. Another issue deals with data policy; data base protection legislation and the continued challenge with restrictions to the full and open access to data. While emerging new technologies are helping us in some of these areas, the issue of the digital divide between haves and have nots is an issue that needs continuous attention.

Finally information must be incorporated from scientific information in addition to federal and non-federal information. The strengths of data and information related to climate at this time involve the following elements:

- A common framework (i.e., community-accepted standards and protocols) for provider's data, data quality control, metadata documentation, and exchange formats that enable public and private sector access and use.

- A policy of free, open and timely access to observational data – nationally and internationally.

- Life cycle observational data management and stewardship (i.e., long term preservation, ongoing quality management, climate record rescue & construction, and knowledge extraction)

- Long-Term Data Sets Emphasis

The goal is an integrated data system in order to better integrate data from different sources to meet the needs of users focusing on 2-4 years and beyond; provide guidelines for formatting the data; implement quality control; facilitate the integration of data from different sources and agencies; and implement life cycle data management for every observation system whose success will rely on national and international partnerships. Fortunately, there are many on-going and solid data management activities already in place that provides a solid foundation for further progress. A few examples include:

- Global Change Master Directory with thousands of pointers to a great diversity of datasets and research work

- The Data Information Working Group (DIWG) that has worked under the auspices of the US Global Change Research Program since 1991 in working to coordinate the distributed data activities of many federal agencies. In conjunction with this, the DIWG's Global Change Data Information Service available on the web at http://globalchange.gov/ and maintained by the Department of Energy (DOE) has been a great data service to the community.

- NASA's Earth System Enterprises which manages a distributed data management process through 8 Data Active Archive Centers (DAAC).

- The NASA Earth System Federation, which manages over 2 petabytes of data serving over 2.5 million customers.

- There are a number of data center activities across the federal Government which include but are not limited to the following:

    - The Carbon Dioxide Information Analysis Center providing a wide variety of data and information products to the global-change community since 1982 (**http://cdiac.ornl.gov).**

    - Various World Data Centers (WDC) hosting a number of different thematic data themes (e.g., meteorology, trace gases, solar terrestrial data) from a number of different federal organizations. These WDCs are part of a larger global system under the auspices of the International Council of Scientific Unions.

    - Data management for AmeriFlux, a network of research sites quantifying flows of carbon between the atmosphere and the terrestrial ecosystems and FACE (Free-Air $CO_2$ Enrichment): large-scale studies of the effects of elevated $CO_2$ on vegetation.

    - An Atmospheric Radiation Monitoring (ARM) archive providing data on atmospheric radiation balance and cloud feedback processes (**www.archive.arm.gov/)**

We will have to continue taking on new challenges; more distributed heterogeneous data management system expanding on past successes. NOAA is experiencing increased data holdings and the growth of its archive is projected to increase from 0.5 petabyte now to over 16 petabytes of data holdings by the year 2015. NOAA is actively involved in planning for this and will be utilizing the concept of Data Stewardship to increase the quality of climate data through a systematic data management program to better monitor system performance on the fly, document, and make data and more importantly products available.

In Chapter 12 on Grand Challenges we are going to have to investigate how are we going integrate all this data; what will the infrastructure be; and what partnerships are necessary

in order to develop a framework for integrating data and determining requirements for the data as well as an active data archaeology program for rescuing and finding new data that will aid in preparing targeted data products under a national data management architecture for climate observing systems.

Next Dr. Roberta Balstad Miller gave her views on the strategic plan and started by indicating the many strengths of the plan particularly in three areas:

- life-cycle data management; decades long time series very important
- open and widespread access to data
- maintenance of global monitoring systems and support for data products

While she did not have specific problems with the plan itself, her comments were directed toward the strategic plan in general. Dr. Miller indicated that we should set up structures to do planning and indicated that long-term strategic plans cannot anticipate everything but rather needs to be flexible to keep it moving ahead and further identified three areas where the CCRI Strategic Plan could be improved:

- First it is too narrowly focused on data accessibility and uncertainty needs of scientists
- Data rescue and reconstruction of past time series is very complex and needs more research particularly regarding areas of socio-economic behavior and institutional data on the forcing and impacts working with other communities
- Finally the system requirements were too vague or in some case non-existent in the plan. Decision makers and non-scientific users are left out of the strategic plan. For example a strength of a Climate Change Information System for decision makers in the NYC area is its emphasis on the needs of these non-science users. First, a survey of the needs of this sector in a form they can use—probably not a scientific data system—is essential. A distinction must be made between scientific uncertainty and needs of knowing how to identify trends must be made to the non-scientific sector.

Data rescue and archeology are unique tasks for the reconstruction of past time series that should be expanded beyond the climate change scientific realm. It needs a new type of science to do this better. The reconstruction of data sets from rescued data requires new skill sets from non-traditional field such as statisticians who understand better than anyone the reconstruction of past time series. Indirect measures of change (e.g., crop studies, soldier nutrition, etc.) are required and need a new kind of researcher. Such socio-economic input is required and the influence of socio-economic factors that are required need to be better incorporated into the strategic plan.

Finally the requirements for data management are not spelled out in the plan; life-cycle management will require a range of professional skills in computer science; and continued innovation will be required to incorporate on the fly new and diverse data sets; array of population data on top of other data. Flexible processes needed to anticipate what inevitably might take place. Long-term data archives will require technological flexibility, and investment in computer science. The Data Management system needs to

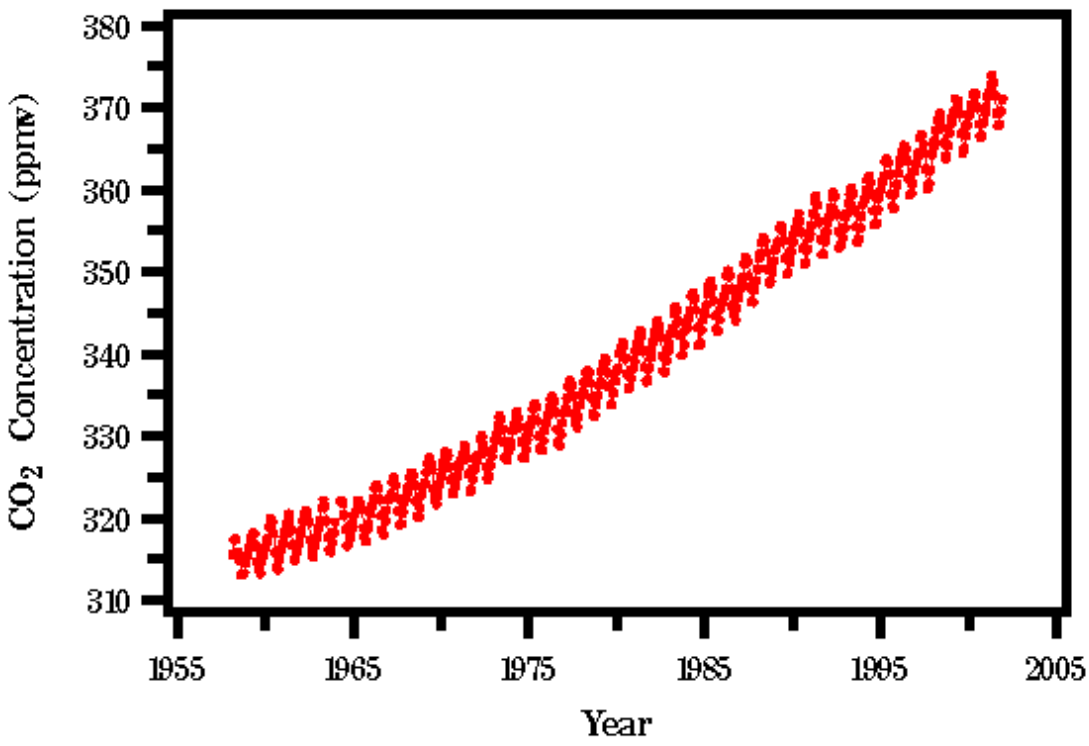
be able to use more on-the-fly integrated data from heterogeneous fields and overlayed upon one another. All aspects of Data Management cannot be predicted now, but planning needs to be flexible to allow for new innovations not yet anticipated.

Greater focus on applied users; research tasks in data archeology; and flexible requirements along with a greater emphasis on research tasks that involve data rescue and archeology will demand a flexible process for evolving data system requirements.

Dr. Tom Boden's reiterated Dr. Miller's view that the plan had a lot of strengths. His view was that the pproposed plan would improve and further global climate-change research and data management. His view was that the proposed plan builds on and transitions nicely from many existing programs, and that the questions, objectives, and priorities discussed are relevant and contemporary with sound scientific basis. The Plan recognizes the need for a life-cycle data management approach and reflects many of the present trends in computing and data management. The CCSP Strategic Plan recognizes that basic climate measurements alone are not adequate to address global climate-change issues, and promotes active collaboration among measurement, modelling, data management, and policy communities.

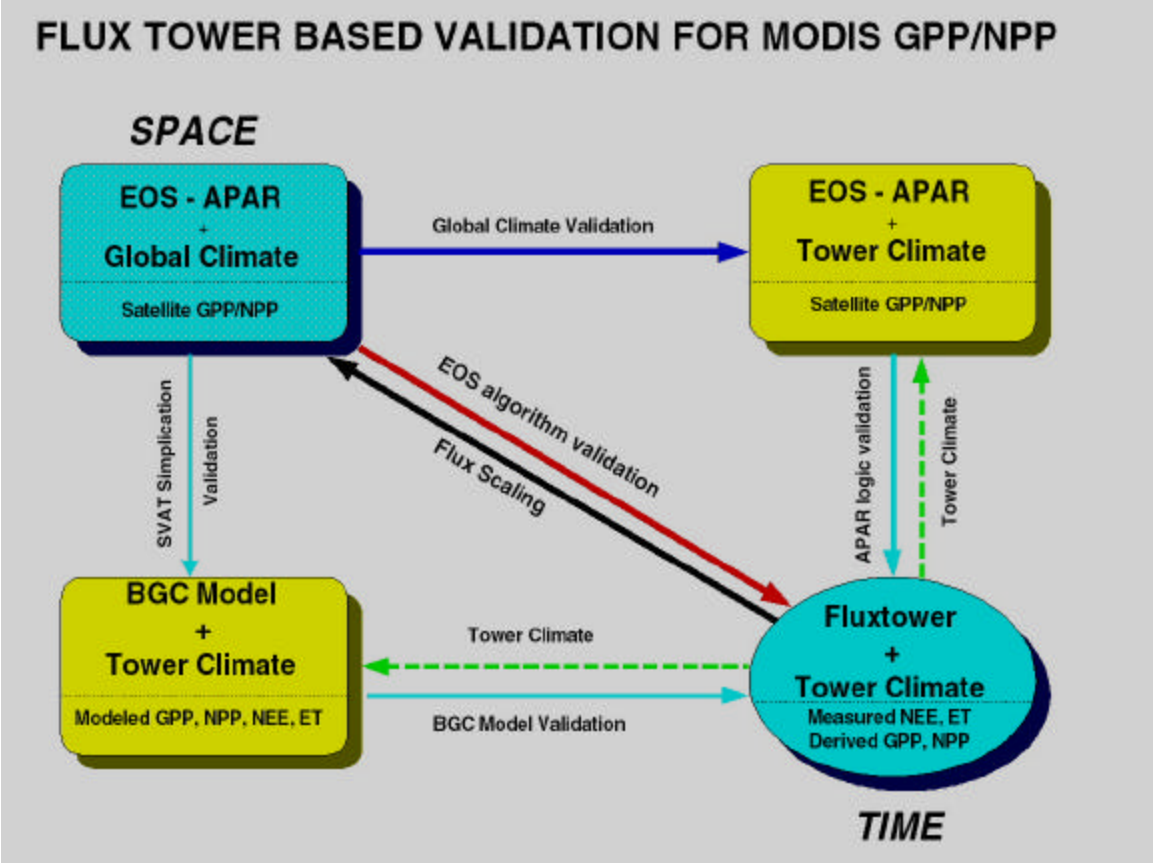
He believed that most key data management issues are discussed in the report, however, some data issues require further attention. Proper focus on oceanic and atmospheric processes but too little coverage of the terrestrial influence on climate was included. The plan identifies the major science and data issues but "strategic plans", by their nature and following on to Dr. Miller's comments, often fail to address how "Research Needs" will be met or "Products and Payoffs" will be accomplished. He generally liked the plan but would now like to see the budget necessary for implementing it.

Regarding observations, a sustained, commitment to long-term monitoring and data information systems is required along with a single coordinated measurement framework to include aerosol measurements, cloud observations, and surface observations is imperative. Long-term data sets such as Dave Keeling's $CO_2$ series (see below) are essential. Coordination and stronger ties to the terrestrial monitoring networks (e.g., LTER, AmeriFlux) is required and new climate simulations with interactive carbon cycle budgets, volcanic and solar ensembles, energy-use impacts, and sulfur cycle linkages (e.g., with varying $SO_2$ emissions) are required.




Source: Dave Keeling and Tim Whorf (Scripps Institution of Oceanography)

On the other hand, gaps in the plan do exist**.** The plan needs stronger linkages to existing data center networks and data projects, and needs to recognize that the "data system" is a distributed WWW-based system adhering to existing XML-based metadata and spatial standards (e.g., FGDC standards). It fails to recognize that "unavailability" of existing data (particularly gridded data) is still an impediment to climate modeling efforts, and needs a greater emphasis on metadata – detecting biases requires supporting documentation. The plan fails to properly address the caveats and limitations of historical data and historical proxy records or the difficulties in documenting these databases, as well as what the data assimilation requirements for "data assimilation modeling are. While he did not read the third part of the plan, Dr. Boden believed the overall plan lacked an education component; according to Dr. Boden, "We must recruit and train future generations of data specialists, micrometeorologists, climate modelers, in order to assure success in this key area of climate data management". He went on to give an example of an exercise that illustrates where we need to go in this area. This was headed by Dr. Steven Running of the University of Montana and involves the measurements of water vapor, carbon dioxide, and energy exchange. The exercise was for AmeriFlux sites to provide meteorological data (e.g., temperature, precipitation, wind, etc.) so that the sites could model the productivity of these sites. The data center provides the data to the modeling groups so that a comparison between actual and modeled data can be made and can use validation for satellite and in-situ data. That exercise is outlined in the following figure below:



In conclusion, Dr. Boden believed it was a good strategic plan that should be implemented now. We must stay the course, as success requires a sustained, long-term commitment to observations, model development, and data information systems. WeMust encourage funding support to strengthen existing data efforts (QA/QC, data rescue, data synthesis, metadata assembly, GIS compilations), encourage continued system development featuring,-WWW-based distributed systems with archival capabilities, automated data processing capabilities, data visualization/analysis capabilities XML-based metadata, GIS coverages, and ready access to ancillary data.

Dr. Lee Dantzler started by talking about the importance of data management to science but we will fail if we do not successfully address the data management issues on the table. He outlined the vision of the future CCSP system as one where the user will easily access and browse across multiple sources of information, be able to compare data in understandable and simply employed scientific graphics using standard web browsers, GIS, and scientific quality visualizations tools without concern for data format, data location, and data volume.

A common framework (i.e., community-accepted standards and protocols) for provider's data, data quality control, metadata documentation, and exchange formats that enable public and private sector access and use is essential as the scientific community is not the primary audience for this information. A policy of free, open and timely access to observational data – nationally and internationally; it is not well stated in the plan and is absolutely critical. Life cycle observational data management and stewardship (i.e., long term preservation, ongoing quality management, climate record rescue & construction, and knowledge extraction), and the plan must address biological/ecosystem observations (especially the oceans). Climate change and variability are critical to the lives of people and that is why we do the data management.

Dr. Dantzler went on to speak to how this data management approach has worked in building the oceanographic climate data record, using the World Ocean Database (2001 version) as a prime example. This database is generally accepted as the most complete compilation of global, scientific, climate data record quality, oceanographic observations in existence for physical, chemical, biological observations, through the use of fixed, towed, free floating, vertically profiling sensor platforms, peer-reviewed quality control, data documentation, and analysis of aggregated database.

Key questions that need to be answered by good data management practices are:

- How has climate changed and with what confidence can natural and human causes be identified?
- What is past, current and future climate state, and how can observations be improved to Support better predictive modeling?
- How to reconcile suggested differences between surface and tropospheric temperature trends?

- How can biological & ecological systems' response to climate change is better understand through improved observations?
- How accessible is the climate record?

To answer these questions and to reiterate what Dr. Dantzler began his talk with we need to do the following:

- Work towards a common framework (i.e., community-accepted standards and protocols) for provider's data, data quality control, metadata documentation, and exchange formats that enable public and private sector access and use.

- Free, open and timely access to observational data, both nationally and internationally.

- Sustained life cycle observational data management and stewardship (i.e., long term preservation, ongoing quality management, climate data record rescue & construction, and knowledge extraction)

- Address biological / ecosystem observational needs

Dr. Richard Rosen indicated that an important tenet of the strategic plan needs to be the continued free, open, and timely exchange of data; this should be a fundamental U.S. policy pushed for internationally as this continues to be a problem internationally especially as this data becomes more economically valuable; access to this data should be as low cost as possible, and he was happy to learn that organizations such as NOAA's National Climatic Data Center were working towards that goal of more affordable access to data holdings. He spoke to the importance of monitoring the climate monitoring system as it is essential to climate monitoring to have high quality systems producing quality data, and if there are problems this monitoring of the system's health was vital to know as soon as possible.

A main concern of Dr. Rosen's is the content of databases more than its eventual distribution. Raw and adjusted measurements and metadata important components, and the placement of adjusted values along with the unadjusted would save the scientific community the effort of learning what went into the databases originally. Recent exercises have shown that agreement in homogeneity of databases is hard to ascertain; the limited agreement between homogeneity adjustments made by different methods shows the difficulty of creating climate-quality radiosonde temperature datasets for example, and he was pleased with the effort of the strategic plan in ensuring the identification of time dependent biases as early as possible; and this remains a major research issue.

Dr. Rosen spoke about the conflicting needs of the weather and climate communities. The Weather community changes systems often in order to take advantage of technology, while the climate community is more conservative and changes much more infrequently in order to maintain homogeneity and minimize transfer functions; while the climate

research community welcomes change, they would do it much less often. Therefore, a natural tension exists between the climate and weather communities that put the National Weather Service (NWS) in a difficult position to balance these issues. While observations can and do serve multiple purposes, it should be noted that as Dr. Jim Anderson stated the previous day, "Climate data are not weather data", and there are different needs in the level of accuracy and homogeneity between climate and weather that need to be considered.

The NWS is currently looking at upgrading its upper air radiosonde system; and in order to reconcile differences, a subset of stations forming a climate quality UA network along with a reference radiosonde for developing measurement standards for these observations are essential elements of maintaining data continuity and homogeneity of the long-term climate data record.

He was pleased with emphasis on Reanalysis in the plan:
*"A strategy for routine reanalysis must be established to exploit the iterative nature of improvements in this process."* [Strategic Plan, p. 27, lines 24-25], but wants to ensure that it incorporates not just atmosphere, but ocean and land as well, and made suitable for trends despite observational heterogeneities. There needs to be a greater emphasis to monitor winds, tropospheric circulation, and transports of energy and water in addition to concentration of plan on precipitation and temperature. A linkage between climate modeling and monitoring is needed to support joint efforts by modeling and diagnostics communities to exploit observations to improve models; communities need to work more closely.

Finally, Dr. Rosen noted the need for an operational agency to speak towards Ten Principles for Climate Monitoring, take ownership of that, and be the voice that speaks for climate quality data in the community.

The summary presentation which was put together for the plenary session and presented by Ms. Maiden later that morning, detailed the vision and strengths of the plan as outlined by Dr. Conkright and others as detailed earlier in this report. In addition a set of suggested additions to the plan that were offered by the various panel members are summarized as follows:

- More fully address terrestrial, biological, ecosystem observations, and social science data
- Establish a structure for setting priorities, for improving the accessibility for both policy and science users, as the plan is too narrowly focused on data accessibility and uncertainty needs of scientists
- Ensure implementation of the 10 Principles of Climate Monitoring
- Data rescue, as highlighted in the section is very complex and needs associated research; climate quality and is something best done up front
- System requirements are too vague in the plan
- Clearer statements on interagency process for implementing
- Elucidate flexible process for community-based standards and protocols

**Question and Answer Session**

Lucia Tsaoussi, NASA – A question on the conflict between climate and weather data was directed to Rick Rosen.

Dr. Rosen's response was that there is not necessarily a conflict, and that climate and weather should not be looked at as being at odds, but it is an issue that needs to be continually looked at.

John Bates, NOAA – What can this community use from other communities (e.g., libraries) to help in getting users to better access and use data. Investigative reporting may be a resource to be used to aid in climate change research re: the construction of unique databases from varied and diverse sources to apply innovative techniques to the climate area.

Roberta Miller responds: Size of the community drives the technology and we need to get away from that in order to get to other disciplines with new databases across the Federal Government as well as the private sector.
Lee Dantzler responds: Not averse to using technology; better linkages of data from varied sources. We do not have to be the drivers for technology and this is probably a blessing.

Bill Rossow, NASA – We are not the experts in technology and we need to get interest from experts from other fields; as such there must be funding with our label on it in order to attract experts from other fields and this must be in the strategic plan. Why do we still talk about RAOBs when that should have been replaced by satellites 10-20 years ago. Current infrastructure is not moving forward at all; and thus Plan must address these requirements. That said, we have to take what we presently have, what the needs are, how they can be implemented in the framework of current stuff knowing that there is no new fountain of money.

Rick Rosen – Sees Upper Air revolution when satellite resolution begins to match RAOB resolution.

Tom Karl, NCDC – How do we move forward in addressing priorities in the Strategic Plan; prioritization is a key but how do we do it?

Roberta Miller – We are frankly not there yet but addressing non-scientific users is a start.

John Townsend, University of Maryland – Getting other communities interested is not just a matter of money, but we need to forge effective partnerships to bring these technologies in the climate arena. Funding opportunities for interdisciplinary opportunities between Computer Science, Libraries, and Climate communities.

Sharon LeDuc, NCDC – No talk at all about emissions and inventories; climate quality data for emissions is essential.

Tom Boden – Data may or may not be there and there are sensitivities regarding release of underlying data. Should be in the plan.
Chris Justice, University of Maryland - What's missing is what agencies are planning over the next 10 years; currently very weak on terrestrial side. For NPOESS moving data to information quickly will be a big help. Roles and responsibilities for transitioning from research to operations and who should do this.

P. Krishna Rao, TRW – Problem is not the data (e.g., MSU data since 1963), the problem is the lack of a climate processing system dedicated to reprocessing the weather data for climate purposes. What is it we can do over the 2-4 years—dedicated climate processing system to reprocess data for climate; this will be a help to NPOESS which is beyond CCSPO.

Lee Dantzler – There are various approaches to reprocessing the data usually left up to community; but what is Climate Processing System; is it a distributed network that is a virtual processing system; does not understand nature of what such a system is.

Rao – The system is total thing including the science and associated agreements.

Rossow – What is missing is a stand-alone system; people manage to eek out reprocessing but on a very small scale. What is missing – (1) products don't come together in any coherent way- regarding Rossow's cloud data there are no mature data products available in a coordinated effort; (2) maturity of methodologies should be handed over to some operational entity; research people do it for a while and then they move on.

Townsend – What are we saying; should there be more coordinated structure; should there be a competitive process for getting records?

Rossow - Must be funding support for researchers to develop methods and then a process to hand it off to an operational entity. Clear funding pots for development and operations to encourage cooperation and coordination.

Greg Withee, NOAA/NESDIS – We would like to do this job better and would take guidance from community.

Miller – Yes we need to set priorities, yes we need operational agencies to undertake priorities, and yes we have to recognize that priorities will have to change over time. Need to take advantage of innovations in statistics sampling, library and computer science, etc. Need most effective and efficient use of resources as possible.

Difficulty of long-range data from researchers

Ed Harrison, NOAA – Uncertainties and helping people to understand these is a challenge. What steps need to be taken to better address this and move this forward in the Strategic Plan.

Gerry Barton – Spoke to the negativity of term uncertainty; and believed that this should be turned around to something more positive like an index of confidence. This idea generated a lot of support from participants who believed that this was a good idea that should be further pursued by the CCSPO.

The breakout session adjourned on time at 1030 hours.